An exon skipping in a *SEPALLATA-Like* gene is associated with perturbed floral and fruits development in cucumber

Summary We isolated a mutant showing perturbations in the development of male and female floral organs and fruits. Analysis of the single nucleotide polymorphisms from bulked F_2 pools identified the causative variant occurring in *Csa4G126690*. *Csa4G126690* shows high homology to *Arabidopsis SEPALLATA2* (*SEP2*) thus being designated *CsSEP2*. The causative variant was located on the splicing site of *CsSEP2*, resulting in the skipping of exon 6 and abolishment of the transcriptional activity. Our data suggest that *CsSEP2* is involved in the floral organ and fruits development by conferring transcriptional activity.

In flowering plants, floral development represents a significant diversion from vegetative growth and is of closely relevant to the yield of plants. The majority of flowering plants produce bisexual flowers with female and male organs on the same plants, while the separation of female and male organs occurs on the same individual for monoecious species and on different plants for dioecious species. *Arabidopsis thaliana* is a model species for bisexual plants, and the ABCE model has been proposed as a key principle to decipher the molecular mechanism of floral specification. According to this model, A + E genes are needed for sepal, A + B + E for petal, B + C + E for stamen, and C + E genes for carpel (Bowman et al. 1989, 1991; Jack et al. 1992; Goto and Meyerowitz 1994; Weigel and Meyerowitz 1994; Riechmann et al. 1996; Pelaz et al. 2000; Ditta et al. 2004; Sridhar et al. 2006). Most of ABCE genes encode MADS-box proteins, suggesting that the MADS-box genes are essential for flower development.

Cucumber (*Cucumis sativus*) is a widely cultivated vegetable crop, and a typical monoecious species with unisexual flowers on the same individual. The formation of unisexual flowers results from the selective arrest of the development of either staminate or pistillate primordia just after the bisexual stage (Malepszy and Niemirowicz-Szczytt 1991; Kater et al. 2001). Two sets of genes are involved in the process of floral development. One set of genes promotes the accurate formation of all organs necessary for a normal flower. The other set of genes determines the sex identity of a flower. At present most studies focus on the isolation of genes responsible for sex determination (Malepszy and Niemirowicz-Szczytt 1991; Li et al. 2012; Boualem et al. 2015), few studies has been performed to investigate the genes involved in the ABCE model in monoecious plants with unisexual flowers on the same individual (Kater et al. 2001).

In this study, we isolated a floral mutant through an ethyl methane sulfonate (EMS) screening in the cucumber inbred line 406 background. Compared with the wild-type plants, the mutant produced enormously large sepals in both female and male flowers (Figure 1A). In normal female flowers three lobes of the stigma were observed, whereas in mutant female flowers three lobes of the stigma became irregularly shaped and were not easily recognized, and the style was nearly three-fold longer than the normal ones (Figure 1B). The male flowers in the mutant were also considerably different and larger than that in the normal plants. In the mutants, the filaments cannot come together, and the inner cavity became filled with a style-like tissue (Figure 1C). We also observed defective fruits in the mutants, as the fruits just grew up to 5 cm and then shed from its peduncle (Figure 1D). To characterize the floral phenotype of the mutant in depth, we conducted microscopic observation of flowers at different development stages (Figure 2A–V). The cucumber floral development has been classified into 12 stages from meristem initiation (stage 1) to anthesis (stage 12), and at stage 6 the male and female flower buds is morphologically distinguishable (Bai et al. 2004). Our results showed no significant difference from stage 1 to stage 7 in both male and female flowers, while at stage 8 morphological differences were visible. Therefore, the abnormal development is likely to occur between stage 7 and stage 8. At stage 8, the pronounced difference was that the carpel primordial elongated in female and male flowers of the mutant plants.

Phenotypic evaluation of a F_2 population derived from a cross between the male mutant and female wild-type plants showed 44 plants bearing aberrant flowers and 182 plants bearing normal flowers. Through χ^2 test, we found that the observed segregation ratio fitted with 3:1 ($P = 0.0548$), indicating that the aberrant floral trait was dominated by a recessive gene.

To identify the variant associated with the phenotype, whole genome re-sequencing was carried out for three pools of DNA, including M pool from 16 F_2 mutant individuals, W pool from 16 F_2 wild-type plants, and 406 pool from 20 background plants. The resultant sequencing data of the three pools were aligned against the cucumber reference genome 9930 (Huang et al. 2009), and the nucleotide sequence was generated for each pool. As a result, 223,473 raw SNP loci for W pool and 229,524 for M pool were extracted for further analyses. Respective 80,960 (36.2%) and 72,325 (31.5%) SNPs in W pool and M pool were eliminated through the filtering process. The remaining SNPs from each pool were put together and created a union set consisting of 176,878 loci. It was expected that the SNP responsible for the recessive phenotype was homozygous in M and 406 pools and heterozygous in W pool. We found 1,153 such SNPs, and then the ones differing between M and 406 pools were retained. Considering the preferential mutations caused by EMS mutagenesis (Drake and Baltz 1976), 30 SNPs exhibiting G to A or C to T mutation were finally kept (Table S1). Among 30 SNPs across seven chromosomes, 23 were located in intergenic regions, while only seven SNPs were located in genic regions (Figure 3A).
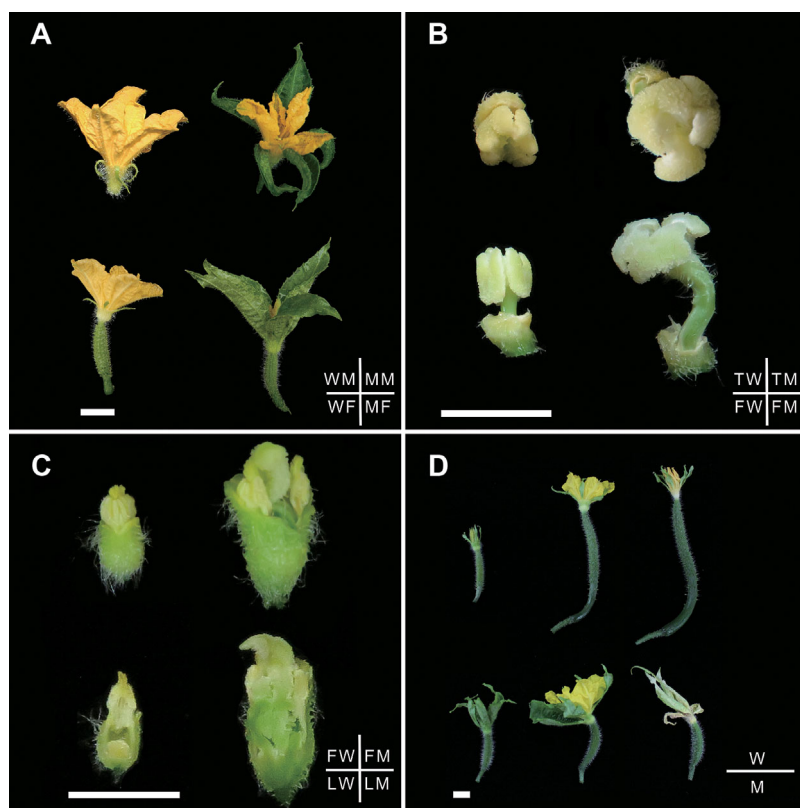
Figure 1. Phenotype of the mutant

Pictures were taken for **(A)** whole flowers, **(B)** style and stigma of the female flowers, **(C)** male flower without sepals and petals, and **(D)** fruits at different developmental periods of the wild-type plants and *Cssep2* mutant. In **(A)**, WM represents wild male flower; MM, mutant male flower; WF, wild female flower; MF, mutant female flower. In **(B)**, TW represents top view of wild type flower; TM, top view of mutant flower; FW, front view of wild type flower; FM, front view of mutant flower. In **(C)**, FW represents front view of wild type flower; FM, front view of mutant flower; LW, longitudinal section of wild type flower; LM, longitudinal section of mutant flower. In **(D)**, W means wild type fruit and M mutant fruit. Scale bar = 1 cm.

Among seven SNPs, two occurred in the intronic regions on chromosome 6. The remaining five SNPs resided on chromosome 4; three of them occurred in intronic regions, and SNP4G12697245 caused nonsynonymous mutation and SNP4G7741482 was present on a splicing site. Amplification of SNP4G12697245 (Table S2) and its flanking region from six mutant individuals of M pool showed that this SNP was heterozygous, so it was excluded for further analysis. A dCAPs marker (Table S2) based on SNP4G7741482 was tested in the F2 population consisting of 213 individuals, to determine the association between the variant and the mutant phenotype. The result showed that SNP4G7741482 completely co-segregated with the phenotype. Based on the facts that analysis of the re-sequencing data delimited the causative SNPs to seven ones and SNP4G7741482 was the only one completely co-segregating with the phenotype, we concluded that SNP4G7741482 was the causative SNP and *Csa4G126990* was the candidate gene.

In *Arabidopsis*, four *SEPALLATA* (*SEP*) genes are E class genes and required for the specification of floral organs. BLAST search showed that *Csa4G126990* exhibited highest homology to *SEP2* in *Arabidopsis*, so it was designated *CsSEP2* hereafter. Alignment of *CsSEP2* with putative *SEP* genes from different species showed that *CsSEP2* lacked the entire MADS-box domain (data not shown), an indispensable domain for *SEP* genes. We examined the RNA-seq data of a previously sequenced cucumber line 9930 (Huang et al. 2009; Li et al. 2011), finding that there were still reads beyond the end of the annotated *CsSEP2* (Figure S1). Those reads supported an extra exon with a length of 185 bp at the 5′ end and a start codon was predicted in the extra exon (Figures 3B, S1A). Together with a single base pair just before the previously annotated start codon (Figure S1B), the extra sequence can be translated into a peptide consisting of 62 amino acids and this short peptide was predicted to contain a MADS-box domain coincidentally. Sequencing of the *CsSEP2* transcripts from wild type plants verified the gene structure of *CsSEP2* (Table S2). SNP4G7741482 was present on the 5′ end of intron 6 (splice donor site), and sequencing of the mutated *CsSEP2* revealed that the whole exon 6 (42 bp) was missing without a frameshift mutation (Figure 3C). Examination of the transcriptomic data of 9930 revealed that *CsSEP2* has a flower-specific expression pattern (Li et al. 2011).

The *Arabidopsis SEP3* has been demonstrated to mediate the interaction between MADS-box proteins by supplying transcriptional activation (Honma and Goto 2001). We performed an assay to test whether the exon skipping event
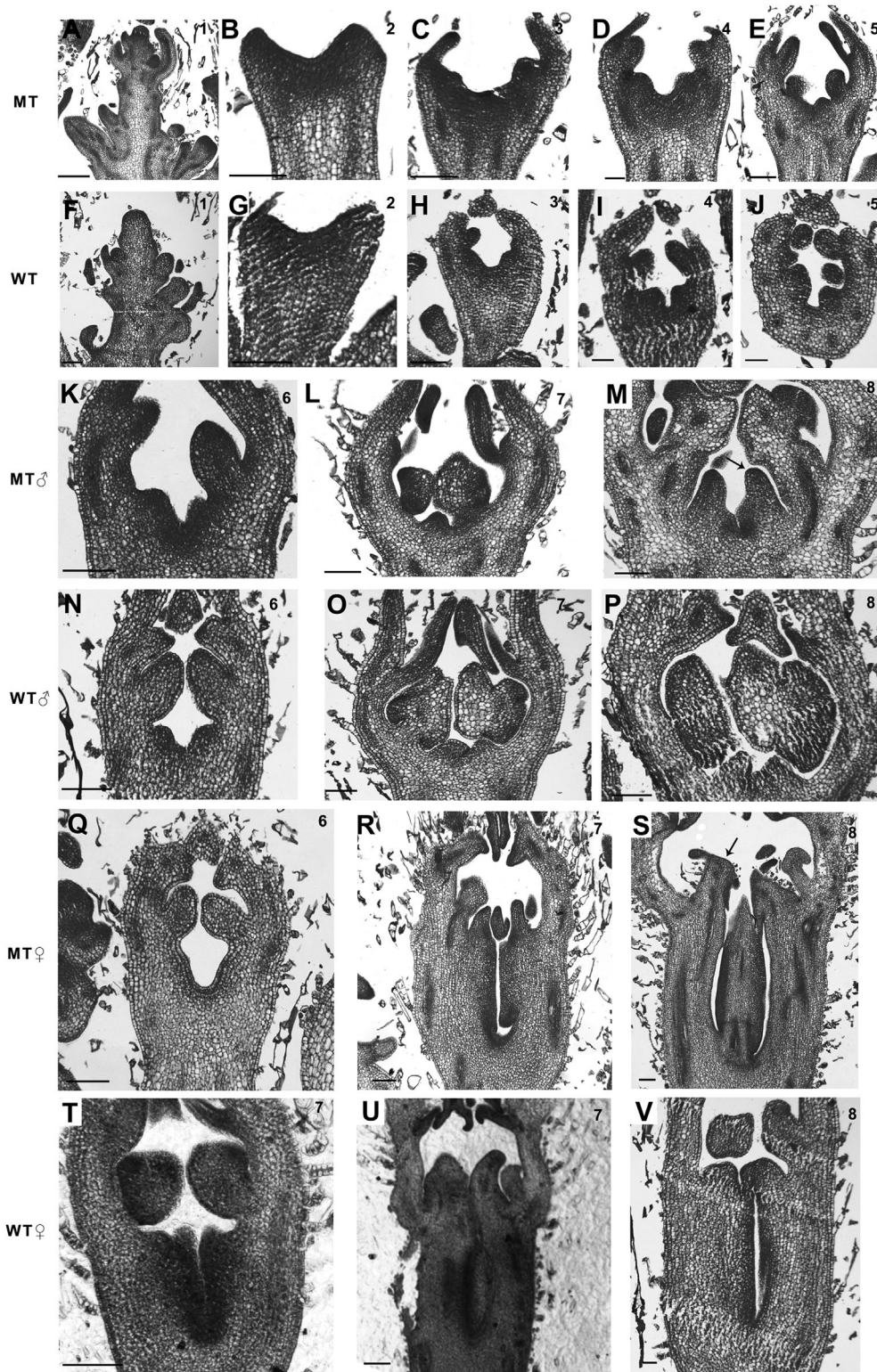
Figure 2. Development of mutant (MT) and wild type (WT) flowers
(A) The longitudinal section of a shoot tip from mutant plants. The floral buds are at early stage 1. (B–E) Floral buds at stages 2 (B), 3 (C), 4 (D) and 5 (E). (F) The longitudinal section of a shoot tip from wild-type plants. The floral buds are at early stage 1. (G–J) Floral buds at stages 2 (G), 3 (H), 4 (I) and 5 (J). (K–M) Mutant male flowers at stages 6 (K), 7 (L) and 8 (M). (N–P) Wild type male flowers at stages 6 (N), 7 (O) and 8 (P). (Q–S) Mutant female flowers at stages 6 (Q), 7 (R) and 8 (S). The arrowheads points the elongated carpel primordial in the female and male flowers of the mutant plants. (T–V) Wild type flowers at stages early 7 (T), 7 (U) and 8 (V). The number in the upper right corner of each picture indicates the developmental stage. Scale bar = 100 μm.
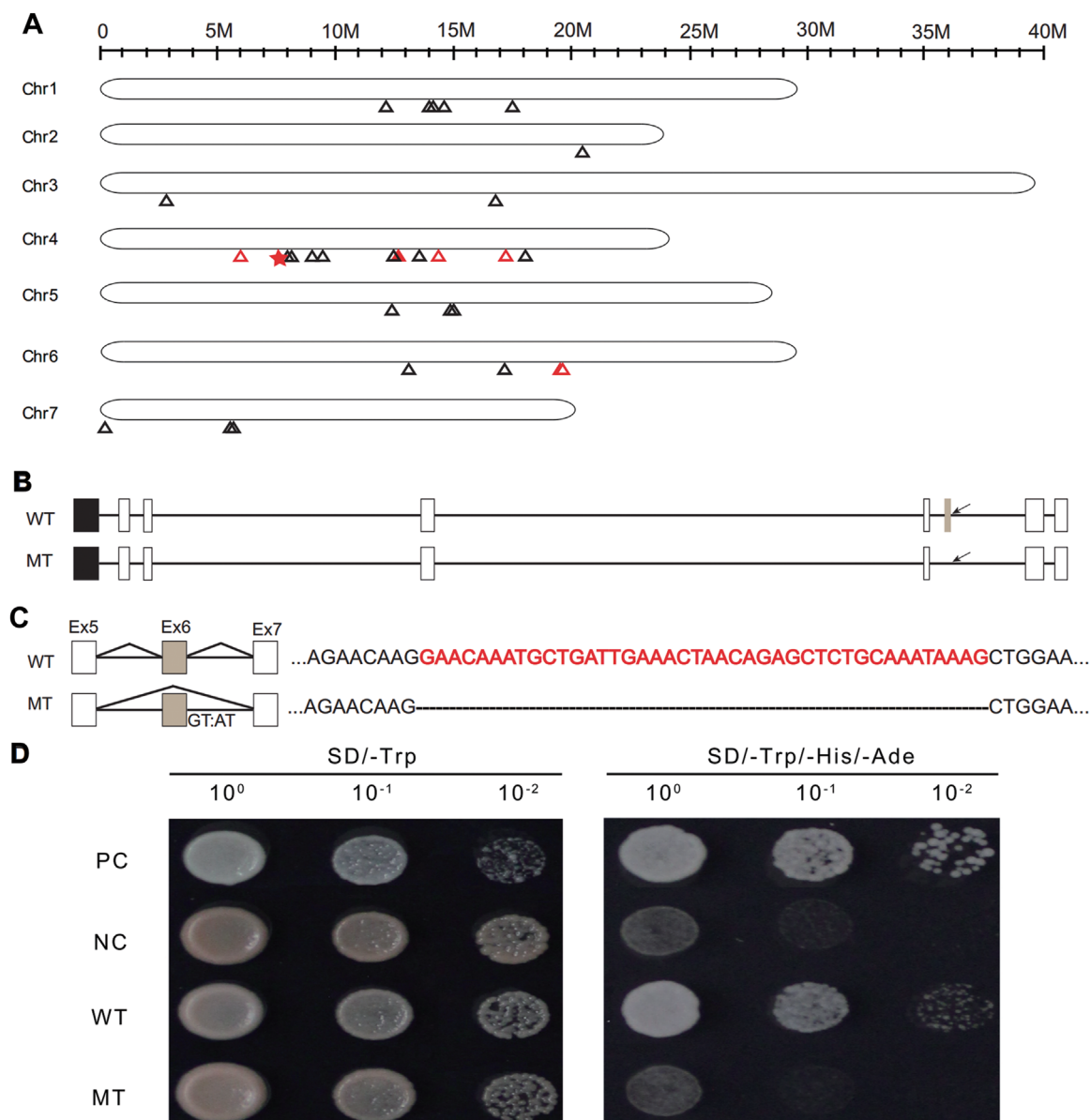
Figure 3. Identification of SNPs associated with the phenotype
(A) Distribution of SNPs after filtering. Black hollow triangles represent 23 SNPs located in intergenic region, red hollow triangles represent five SNPs located in the intronic region, red solid triangle indicates the non-synonymous mutation and red five-pointed star indicates the SNP in the splicing site. **(B)** Gene structure of *CsSEP2* in the wild-type (WT) plants and mutant (MT). Lines indicate the intron and boxes stands for exon. Black solid box indicates the first exon which was inferred from RNA-seq data and validated by PCR. Grey solid box indicates exon 6 in which the causative SNP was detected. **(C)** Exon 6 was spliced and 42 nucleotides (red letter) were missing. **(D)** Transcriptional activity assay in yeast. Yeast harboring pGBKT7-VP16 (PC) and pGBKT7-*CsSEP2* (WT) grew on both SD/-Trp and SD/-Trp/-His/-Ade media while the yeast harboring the empty pGBKT7 (NC) or pGBKT7-*CsSEP2* (MT) only grew on SD/-Trp media.

affected the transcriptional activity of *CsSEP2*. The yeast transformed by the construct containing the coding region of wild type *CsSEP2* grew on the SD/-Trp/-His/-Ade plates, in contrast, the yeast transformed by the construct containing the coding region of mutant *CsSEP2* could not grow (Figure 3D; Table S2). These results indicated that the exon skipping event originating from the SNP on the splicing site abolished the transcriptional activity of *CsSEP2*.

Mutants are valuable sources to uncover the gene for a trait of interest, and the utilization of advanced analysis tools expedites the process. In this study, we applied modified MutMap (Abe et al. 2012) to localize the gene controlling the phenotype. Analysis of the SNPs from three bulks identified the causative variant occurring in *CsSEP2*. In the mutant, exon 6 was spliced out because of the presence of a SNP on the splicing site of exon 6. The RNA splicing mechanisms involving

the splicesome consisting of five uridine-rich small ribonucleoproteins (UsnRNPs) have been elucidated (Staley and Guthrie 1998; Wahl et al. 2009). In the case of *CsSEP2*, snRNP U2 bound to 5′ end of exon 6 searches for the adjacent snRNPU1 which was supposed to bound to 3′ end of exon 6, however, snRNP U1 could not recognize 3′ end of exon 6 owning to the mutation. Thereby snRNP U2 may interact with snRNP U1 bound to 3′ end of exon 7, which would produce an exon (exon 6 + intermediate intron + exon 7) longer than the defined exon (<300 bp) (De Conti et al. 2013), therefore the splicesome tends to skip the exon 6 to create a short mRNA.

Using corrected *CsSEP2* amino acid sequence as a query to perform BLAST against cucumber genome, two genes, *Csa6G095270* and *Csa1G039900*, with homology to *CsSEP2* were found. Similar to the case of *CsSEP2*, these two genes were also erroneously annotated in 9930 and lacked the MADS-box domain. We corrected the nucleotide sequences of these two genes in the same manner as for *CsSEP2*. Alignment of these three putative *SEP* genes showed 49.8% identity, whereas four *SEP* genes in *Arabidopsis* showed 41.2% identity (Figure S2). In *Arabidopsis*, disruption of a single *SEP* caused only subtle phenotype (Pelaz et al. 2000; Ditta et al. 2004), as opposed to the severe perturbation of floral development in cucumber by mutation of a single *CsSEP2*. Why *CsSEP2* plays such an important role merits further investigation.

The *SEP* genes play a broad and crucial role in flower development in *Arabidopsis*. The primary role of SEP proteins seems to provide a scaffold for the formation of transcription complexes and convey upon these complexes transcriptional activation potential (Goto et al. 2001; Honma and Goto 2001). In this study, we found that the exon skipping event abolished the transcriptional activity of *CsSEP2* (Figure 3D), which means that the primary role of *CsSEP2* was severely disrupted and suggests that the complexes of MADS-box proteins could not execute their proper functions. Since pollen from the mutant can be used for crossing, the stamen may generally function as normal ones although it possesses considerable morphological variation. Petals in mutant plants appeared the same as in wild type plants. In contrast, the sepal and stigma were considerably different from normal ones. These morphological observations led us to speculate that the interaction between malfunctional *CsSEP2* and other MADS-box proteins may be disproportionately affected. It has been proposed that tomato *SEPALLATA* homologs, *TM5* and *TM29* play a role in fruit development (Pnueli et al. 1994; Ampomah-Dwamena et al. 2002). We observed that fruits of the mutant cannot elongate, suggesting that *CsSEP2* may also function in fruit development (Pnueli et al. 1994; Ampomah-Dwamena et al. 2002). Further studies identifying *CsSEP2*-interacting genes and investigating the influence of the exon skipping event on interactions will help unravel the mechanisms controlling floral and fruits development in unisexual plants.

ACKNOWLEDGEMENTS

Xin Wang[1,2†], Dongli Gao[2,3†], Jinjing Sun[2], Min Liu[2], Yaoyao Lun[2], Jianshu Zheng[3], Shenhao Wang[2], Qingzhi Cui[4], Xiaofeng Wang[1*] and Sanwen Huang[1,2*]
[1]State Key Laboratory of Crop Stress Biology for Arid Areas, College of Horticulture, Northwest A&F University, Yangling 712100, China
[2]Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[3]Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China
[4]Vegetable Research Institute, Hunan Academy of Agricultural Sciences, Changsha 410125, China
[†]These authors contributed equally to the work.
*Correspondences: wangxff99@nwsuaf.edu.cn; huangsanwen@caas.cn

AUTHOR CONTRIBUTIONS

REFERENCES

Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S, Terauchi R (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. **Nat Biotechnol** 30: 174–178

Ampomah-Dwamena C, Morris BA, Sutherland P, Veit B, Yao JL (2002) Down-regulation of TM29, a tomato *SEPALLATA* homolog, causes parthenocarpic fruit development and floral reversion. **Plant Physiol** 130: 605–617

Bai SL, Peng YB, Cui JX, Gu HT, Xu LY, Li YQ, Xu ZH, Bai SN (2004) Developmental analyses reveal early arrests of the spore-bearing parts of reproductive organs in unisexual flowers of cucumber (*Cucumis sativus* L.). **Planta** 220: 230–240

Boualem A, Troadec C, Camps C, Lemhemdi A, Morin H, Sari MA, Fraenkel-Zagouri R, Kovalski I, Dogimont C, Perl-Treves R (2015) A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. **Science** 350: 688–691

Bowman JL, Smyth DR, Meyerowitz EM (1989) Genes directing flower development in *Arabidopsis*. **Plant Cell** 1: 37–52

Bowman JL, Smyth DR, Meyerowitz EM (1991) Genetic interactions among floral homeotic genes of *Arabidopsis*. **Development** 112: 1–20

De Conti L, Baralle M, Buratti E (2013) Exon and intron definition in pre-mRNA splicing. **Wiley Interdiscip Rev RNA** 4: 49–60

Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF (2004) The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. **Curr Biol** 14: 1935–1940

Drake JW, Baltz RH (1976) The biochemistry of mutagenesis. **Annu Rev Biochem** 45: 11–37

Goto K, Kyozuka J, Bowman JL (2001) Turning floral organs into leaves, leaves into floral organs. **Curr Opin Genet Dev** 11: 449–456

Goto K, Meyerowitz EM (1994) Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. **Gene Dev** 8: 1548–1560

Honma T, Goto K (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. **Nature** 409: 525–529

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P (2009) The genome of the cucumber, *Cucumis sativus* L. **Nat Genet** 41: 1275–1281

Jack T, Brockman LL, Meyerowitz EM (1992) The homeotic gene APETALA3 of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. **Cell** 68: 683–697

Kater MM, Franken J, Carney KJ, Colombo L, Angenent GC (2001) Sex determination in the monoecious species cucumber is confined to specific floral whorls. **Plant Cell** 13: 481–493

Li Z, Wang S, Tao Q, Pan J, Si L, Gong Z, Cai R (2012) A putative positive feedback regulation mechanism in *CsACS2* expression suggests a modified model for sex determination in cucumber (*Cucumis sativus* L.). **J Exp Bot** 4475–4484

Li Z, Zhang Z, Yan P, Huang S, Fei Z, Lin K (2011) RNA-Seq improves annotation of protein-coding genes in the cucumber genome. **BMC Genomics** 12: 540

Malepszy S, Niemirowicz-Szczytt K (1991) Sex determination in cucumber (*Cucumis sativus*) as a model system for molecular biology. **Plant Sci** 80: 39–47

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. **Nature** 405: 200–203

Pnueli L, Hareven D, Broday L, Hurwitz C, Lifschitz E (1994) The *TM5* MADS box gene mediates organ differentiation in the three inner whorls of tomato flowers. **Plant Cell** 6: 175–186

Riechmann JL, Krizek BA, Meyerowitz EM (1996) Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. **Proc Natl Acad Sci USA** 93: 4793–4798

Sridhar VV, Surendrarao A, Liu Z (2006) *APETALA1* and *SEPALLATA3* interact with *SEUSS* to mediate transcription repression during flower development. **Development** 133: 3159–3166

Staley JP, Guthrie C (1998) Mechanical devices of the spliceosome: Motors, clocks, springs, and things. **Cell** 92: 315–326

Wahl MC, Will CL, Lührmann R (2009) The spliceosome: Design principles of a dynamic RNP machine. **Cell** 136: 701–718

Weigel D, Meyerowitz EM (1994) The ABCs of floral homeotic genes. **Cell** 78: 203–209

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Figure S1. RNA-seq alignment of upstream regions of previously annotated *CsSEP2*

(A) Grey rectangles indicate normal paired end reads; rectangles in other colors indicate mapped paired end reads with abnormal insert sizes or mapping orientations; the original and modified gene models of *CsSEP2* are displayed at the bottom. **(B)** An enlarged view of the first exon in original annotation. The red box indicates the position of the start codon on the minus strand.

Figure S2. Respective alignment of *SEP* genes in Arabidopsis and cucumber

(A) Four *SEP* genes were aligned in *Arabidopsis*. **(B)**Three putative *SEP* genes were aligned in cucumber.

Table S1. Detailed SNP information

Table S2. Primers used in this study

Supplemental plant materials and methods